Chromosome-level genome assembly and whole-genome resequencing of topmouth culter (*Culter alburnus*) provide insights into the intraspecific variation of its semi-buoyant and adhesive eggs

Haifeng Jiang[1], Yuting Qian[2], Zhi Zhang[3], Minghui Meng[4], Yu Deng[5,6], Gaoxue Wang[1], Shunping He[2*], Liandong Yang[2*],

[1]*College of Animal Science and Technology, Northwest A&F University, Xinong Road 22nd, Yangling, Shaanxi 712100, China*

[2]*State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, 430072, China*

[3]*Fujian Key Laboratory on Conservation and Sustainable Utilization of Marine Biodiversity, Fuzhou Institute of Oceanography, Minjiang University, Fuzhou, 350108, China*

[4]*Key Laboratory of Horticultural Plant Biology (MOE), College of Horticulture and Forestry Sciences, Huazhong Agricultural University, Wuhan 430070, China.*

[5]*State Key Laboratory of Developmental Biology of Freshwater Fish, Hunan Normal University, Changsha 410081*

[6]*Life Science College, Hunan Normal University, Changsha 410081, China*

*Correspondence author: Shunping He (heshunpingihb@163.com); Liandong Yang (yangliandong1987@163.com).

Haifeng Jiang and Yutiang Qian contributed equally to this work.

Abstract

Topmouth culter (*Culter alburnus*) is an ecologically and economically important species belonging to the subfamily Culterinae that is native to and widespread in East Asia. Intraspecific variation of semi-buoyant and adhesive eggs in topmouth culter provides an ideal opportunity to investigate the genetic mechanisms of spawning habits underlying the adaptive radiation of cyprinids in East Asia. In this study, we present a chromosome-level genome assembly of topmouth culter and re-sequenced 158 individuals from six locations in China covering three geographical groups and two egg type variations. The topmouth culter genome size was 1.05 Gb, with a contig N50 length of 17.8 Mb and anchored onto 24 chromosomes. Phylogenetic analysis showed that the divergence time of the Culterinae was coinciding with the time of initiation of the Asian monsoon intensification. Gene family evolutionary analysis indicated that the expanded gene families in topmouth culter were associated with dietary adaptation. Population-level genetic analysis indicated clear differentiation among the six populations, which were clustered into three distinct clusters, consistent with their geographical divergence. The historical effective population size of topmouth culter correlated with the Tibetan Plateau uplifting according to the demographic history reconstruction. A selective sweep analysis between adhesive and semi-buoyant egg populations revealed the genes associated with the hydration and adhesiveness of eggs, indicating divergent selection toward different hydrological environments. The present study offers a high-resolution genetic resource for further studies on evolutionary adaptation, genetic breeding, and conservation of topmouth culter, providing insights

into the molecular mechanisms for egg type variation of East Asian cyprinids.

KEYWORDS

Culter alburnus, population genomics, genetic diversity, egg type variation

1 Introduction

Speciation and ecological adaptations in endemic species are important concepts in evolutionary biology. Cyprinidae (Teleostei: Cypriniformes) is a species-rich family of freshwater fish comprising approximately 3,000 species and 367 genera (Nelson et al., 2016). This diverse group of cypriniformes is distributed widely in Europe, Asia, Africa, and North America. The rapid burst of speciation in cyprinid fishes in East Asia has been suggested to be related to the Qinghai-Tibet Plateau (QTP) uplifting and the Asian monsoon climate formation, which resulted in a cross-linked, river–lake system, leading to remarkable ecological and phenotypic diversification in morphologies, feeding habits, life histories, and reproductive strategies (Chen, 1998; Feng et al., 2022; He et al., 2004). The endemic East Asian cyprinids derived from a single clade of 154 species (Nelson et al., 2016; Tan & Armbruster, 2018; Wang et al., 2007) represent a useful model for investigating rapid speciation and adaptive radiations. However, limited genomic resources are available for this evolutionary important group, and the genomic variations to identify diverse adaptive radiations and speciation have been reported in a few species (Jian et al., 2021; Wang et al., 2015; Xu et al., 2019).

Topmouth culter (*Culter alburnus*) is a Culterinae fish species belonging to the

family Cyprinidae, and in China, it is known as the white fish with high economic value (Ren et al., 2019). It is a widespread species distributed across major drainages in China, except Tibetan Plateau, inhabiting divergent habitats such as rivers, reservoirs, and lakes (Qi et al., 2015; Sun et al., 2021). Interestingly, the topmouth culter has two ecotypes that differ in spawning habits: in lakes such as Liangzi Lake and Taihu Lake in the Yangtze River basin, it lays adhesive eggs that stick to aquatic plants or rocks, whereas in other water bodies (including both rivers and lakes), it spawns semi-buoyant (floating) eggs that float in fast-flowing waters (Chen et al., 2022; Sun et al., 2021) Evolutionary reconstruction of the endemic East Asian cyprinids has revealed an ancestral state of spawning adhesive egg and subsequent independent parallel evolution of floating eggs, in which some species of cultrins and xenocyprinins evolved a transition from floating eggs to adhesive eggs again, such as that found in the topmouth culter (Chen et al., 2021; Chen et al., 2023b; Cheng et al., 2022). Notably, the differentiation of adhesive and semi-buoyant eggs has been suggested as an adaptation for lentic and lotic habitats, respectively (Chen et al., 2021; Chen et al., 2023a). This adaptation is closely related to the cross-linked river–lake system shaped by the East Asian monsoon climate during the middle Miocene, which drove the adaptive radiation of the endemic East Asian cyprinids (Chen et al., 2022; Chen, 1998; Cheng et al., 2022). Therefore, the intraspecific variation of egg types in topmouth culter is conducive to the study of genetic mechanisms of ecological adaptation of spawning habits in the East Asian cyprinids. Studies of the egg types of topmouth culter are limited to the embryonic development, and a recent study by Chen et al., (2022) revealed several key

89 pathways associated with egg hydration and adhesiveness in the embryonic
90 development of floating and adhesive eggs of topmouth culter through transcriptomic
91 and proteomic analyses (Chen et al., 2022). However, the genomic variations in
92 different geographic populations and ecotypes remain to be elucidated.

93 In recent decades, topmouth culter has become an important aquaculture species
94 in China owing to its delicious taste and high economic value. Hybrid lineages of
95 topmouth culter with *Megalobrama amblycephala* (Ren et al., 2019) and
96 *Ancherythroculter nigrocauda* (Zhang et al., 2020) exhibiting desirable traits have been
97 developed. However, overfishing, water pollution, and habitat fragmentation or loss
98 have been threatening the natural populations of topmouth culter (Qi et al., 2015).
99 Previous studies based on putatively neutral markers such as mitochondrial DNA and
100 microsatellites have reported the genetic structure of the geographic isolated
101 populations of the species (Qi et al., 2015; Sun et al., 2021; Xiong et al., 2019). All
102 these studies indicate that wild topmouth culter resources must be protected to prevent
103 further decline of its populations. Genetics analysis based on the mitochondrial DNA
104 control region suggests that a population in the Pearl River basin is distinct from those
105 in the Yangtze River and Heilongjiang River basins, potentially existing as a cryptic
106 subspecies (Xiong et al., 2019). A study employing the mitochondrial D-loop region
107 and microsatellites reported that two semi-buoyant egg spawning populations in the
108 Xingkai Lake in Amur River (Heilongjiang River) basin and the Danjiangkou Reservoir
109 in Yangtze River basin were genetically distant from the other four studied populations
110 (Sun et al., 2021). However, traditional genetic methods are less efficient in revealing

the fine-scale genetic structure, evolutionary history, genomic signature, and key adaptive loci related to local adaptation and egg type variations of topmouth culter. Thus, highly efficient population genomic approaches that can provide more information on genetic parameters, adaptation, and diversification are required.

High-quality genomic data are essential for investigating the genomic variations, genetic diversity, and demographic history of topmouth culter. The first genome assembly of topmouth culter was constructed using short-read Illumina and long-read PacBio sequencing, covering 1.02 Gb in 5,742 scaffolds, with a contig N50 length of 72.24 kb (Ren et al., 2019). However, a chromosomal-level genome assembly of topmouth culter is not yet available. In the present study, an improved chromosome-scale genome assembly of topmouth culter was obtained by a combination of Illumina, PacBio, and Hi-C scaffolding techniques. In addition, a population genetic analysis was performed to investigate the population structure and gain insights into the intraspecific variation of egg types in topmouth culter. The high-quality chromosome-level genome assembly can serve as a valuable genetic resource not only for molecular breeding and conservation of topmouth culter but also for further investigations of ecological speciation in evolutionary radiation of endemic East Asian cyprinids.

2 MATERIALS AND METHODS

2.1 Ethics statement, sample collection, and genome sequencing

All the procedures were conducted following the Animal Care and Ethics Regulations of the Animal Experiment Committee (DK2021030), Northwest A&F University

(Yangling, China). For genome sequencing, a wild female topmouth culter individual captured from the Xingkai Lake (45.2685 °N, 132.7964 °E) in Heilongjiang Province, China, in September 2021 was used. Genomic DNA was extracted from the muscle tissues by using the DNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. To assist the genome annotation, total RNAs of seven tissues (heart, liver, brain, spleen, kidney, muscle, and ovary) were extracted using an EZNA HP Total RNA Kit (Omega Bio-tek) and pooled together for cDNA library construction. Paired-end libraries with an insert size of 300 bp were constructed and sequenced on the MGISEQ-T7, a new sequencer from MGISEQ platform launched by the Beijing Genomics Institute (BGI) Tech based on DNA nanoball technology. The MGISEQ platform promises to deliver high-quality sequencing data faster at lower prices and its performance has been demonstrated to be comparable with the Illumina platform in various studies, including whole-genome, whole-exome, transcriptome, single-cell transcriptome and metagenome sequencing (Zhu et al., 2021). For PacBio library construction and sequencing, high-quality DNA was subjected to size selection by using the BluePippin system, and ~20-kb SMARTbell libraries were prepared and ran on the PacBio Sequel II CLR platform (PacBio Biosciences). Hi-C libraries were prepared and sequenced on the MGISEQ-T7 platform for the chromosome-level genome assembly.

A total of 158 individuals of topmouth culter were collected. These individuals belonged to six populations, namely Xingkai Lake (XKL; n = 30) in the Amur River basin (Heilongjiang basin); Dangjiangkou Reservoir (DJKR; n = 38), Yuanshui River (YSR; n = 22), Liangzi Lake (LZL; n = 17), and Taihu Lake (TL; n = 22) in the Yangtze

River basin; and Hanjiang River (HJR; n = 29) in the Peral River basin. Among these populations, LZL and TL populations lay adhesive eggs, whereas the other four populations (DJKR, YSR, XKL and HJR) spawn semi-buoyant eggs. The caudal fin tissues used for genomic DNA extraction were preserved in 100% ethanol at 4 °C.

2.2 Genome survey and de novo assembly

Filtered MGISEQ-T7 sequencing data were used to estimate the topmouth culter genome size by using a 17 k-mer depth frequency distribution analysis. A total of 48,598,039,979 k-mers with the expected depth value of 46.01 were obtained (Figure S1). The genome size was calculated as the ratio of k-mer number and k-mer depth. Adapter and low-quality regions of PacBio long reads were removed to obtain subreads. NextDenove v. 2.3 (Chin et al., 2016) was used to choose different parameters for multiple assembly versions, and the final assembly with the default parameters was chosen. For genome correction, the PacBio reads and MGISEQ-T7 reads were aligned to the assembled genome and NextPolish v. 1.3.1 (Hu et al., 2020) was employed to polish the initial genome. For chromosome-level assembly, Bowtie2 v. 2.2.5 (Langmead & Salzberg, 2012) was used to align the clean Hi-C data to the assembled contigs, which were further used to construct the inter/intra chromosomal contact map by using Hic-Pro v. 2.11 (Servant et al., 2015). The valid interaction pairs were further ordered, oriented, and anchored to the 24 pseudochromosomes with Lachesis (Burton et al., 2013) by using an agglomerative hierarchical clustering method.

To evaluate the quality of the topmouth culter genome, the MGISEQ-T7 short reads and RNA-seq data were aligned to the assembly, and the mapping ratio and

coverage were assessed. Finally, the genome completeness was evaluated by BUSCO v. 5.2.0 using the Actinopterygii gene set (Simão et al., 2015).

2.3 Genome annotation

Simple sequence repeats (SSRs) in the topmouth culter genome were identified by MISA (Thiel et al., 2003) using default parameters. For other repetitive sequences, the Extensive de novo TE Annotator pipeline was used to identify transposable elements (Ou et al., 2019).

Protein-coding genes were predicted based on homology and de novo and RNA-seq-based strategies. For RNA-seq-based prediction, transcriptome data of liver, heart, kidney, muscle and brain tissues were aligned to the topmouth culter genome and then used for gene structure prediction by using PASA v. 2.0.2 (Haas et al., 2003). For de novo prediction, the high-quality data set generated using PASA was utilized to train *ab initio* gene predictors including Augustus, SNAP, GlimmHmm, and Geneid. For homology-based prediction, protein sequences from seven species were mapped to the topmouth culter genome and then homologous genes were predicted using GeMoMa (Keilwagen et al., 2016). Finally, all gene models were integrated using EVidenceModeler (EVM) (Haas et al., 2008) to generate a nonredundant gene set, and transponPSI (Yagi et al., 2014) was used to remove the genes with transposons. Functional annotation of the translated amino acid sequences of the final gene sets was conducted by alignment to the known databases including Non-Redundant Protein Sequence Database (NR), Gene Ontology (GO), InterPro and Kyoto Encyclopedia of Genes and Genomes (KEGG) by using BlastP with an *E*-value threshold of 1e-05.

2.4 Gene family and phylogenomic analysis

To determine gene family evolution in the topmouth culter genome, orthologous and paralogous gene families were clustered using the gene models from 11 species, namely *Oryzias latipes*, *Triplophysa tibetana*, *Triplophysa dalaica*, *Danio rerio*, *Ancherythroculter nigrocauda*, *Megalobrama amblycephala*, *Ctenopharyngodon idella*, *Hypophthalmichthys nobilis*, *Paracanthobrama guichenoti*, *Onychostoma macrolepis*, and *C. alburnus*, by OrthoMCL (Li et al., 2003). Gene family expansion or contraction was estimated by comparing the cluster size between the ancestor and each species by using CAFÉ (De Bie et al., 2006). GO and KEGG enrichment analyses were performed for expanded and contracted gene families by using the Fisher's exact test. The single-copy orthologous genes were used for the phylogenetic analysis and divergence time estimations. Multiple sequence alignment was conducted using MAFFT v. 7.429 (Katoh & Standley, 2013), and poorly aligned regions were filtered and removed using Gblocks v. 0.91b (Castresana, 2000). Phylogenetic relationships were inferred using RAxML v. 1.5 (Silvestro & Michalak, 2012), with medaka as the outgroup species. The MCMCtree program implemented in the PAML software package (Yang, 2007) was used to estimate the divergence times. Seven calibration time points retrieved from the TimeTree database were applied in the current study (Kumar et al., 2017).

2.5 Single nucleotide polymorphism calling and filtering

Genomic DNA of 158 individuals collected from six geographical populations was used to construct libraries with an average insert size of 300 bp and then sequenced using

the MGISEQ-T7 platform. Raw reads containing adaptor sequences, poly-N (>10%), and low-quality bases (Phred quality value <15) were removed. High-quality clean data were mapped to the topmouth culter genome by using BWA v. 0.7.17 (mem –M –t 20 -k 32). SAMtools v. 1.9 (Li et al., 2009) was used to filter duplicate and unmapped reads, sort the reads, and convert them into the BAM format. Single nucleotide polymorphisms (SNPs) and insertions and deletions (InDels) were identified using the HaplotypeCaller module in GATK v. 4 (McKenna et al., 2010). Raw variant dataset without high confidence were filtered using VariantFiltration in GATK with the parameters "-- filterExpression --Quality (QUAL) < 30.0, QualByDepth (QD) < 2.0, FisherStrand (FS) > 60.0, RMSMappingQuality (MQ) < 40.0, StrandOddsRatio (SOR) > 4.0, MappingQualityRankSumTest (MQRankSum<-12.5), and ReadPosRankSum (RPRankSum <-8.0)". The genomic variants for population analysis were further filtered using VCFTOOLS v. 0.1.13 (Danecek et al., 2011) with the parameters "--min-alleles 2 --max-alleles 2 --min-meanDP 5 --maf 0.05 --max-missing 0.5." Finally, SNPs and InDels were annotated to their corresponding chromosomal locations.

2.6 Population genetic analysis

Based on the genome-wide SNPs and InDels, Plink v. 1.9 (Chang et al., 2015) was used to remove the SNPs which has a high linkage disequilibrium level (--indep-pairwise 100kb 1 0.8). Principal component analysis (PCA) was performed using EIGENSOFT v. 6.14 (Patterson et al., 2006). Population structure was further inferred by using ADMIXTURE v. 1.3.0 (Alexander & Lange, 2011) without prior population information (*K* ranges from 2 to 10), and 10-fold cross-validation was performed to

determine the probable number of ancestors. Neighbor-joining (NJ) tree was generated through IQ-TREE v. 2 (Minh et al., 2020) using the ultrafast bootstrap approach with 1000 replicates. To infer the historical changes in effective population sizes in response to climatic change, we selected one individual from each population with the highest sequencing depth and employed the pairwise sequentially Markovian coalescent (PSMC) (Li & Durbin, 2011) method with a mutation rate (μ) of 4×10^7 and an estimated time of 3 years per generation. The uplift process of the Tibetan Plateau and the time range of three phases of intense uplift (Qingzang, Kunhuang and Gonghe Movement) were predicted based on previous studies (An et al., 2001; Li & Fang, 1999). Genetic diversity indexes, including observed heterozygosity (H_O), expected heterozygosity (H_E), and nucleotide diversity (π), were estimated using populations in Stacks, and population genetic differentiation index (F_{ST}) were calculated using the VCFTOOLS to estimate both global and pairwise divergence among populations. Linkage disequilibrium (LD) analysis for each population was conducted on the basis of the coefficient of determination (r^2) between two given SNPs by using POPLDDECAY (https://github.com/BGI-shenzhen/PopLDdecay).

2.7 Genomic variation analysis

To identify different loci potentially influencing intraspecific variation in topmouth culter egg types, we conducted genomic selective sweeps analysis between two pairwise groups: (a) adhesive egg populations (LZL and TL) and floating egg populations (DJKR, HJR, XKL, and YSR). Nucleotide diversity (π) ratio and divergence index (F_{ST}) were estimated using VCFTOOLS with a 200-kb sliding

window in 20-kb steps. D_{xy}, an absolute measure of genetic divergence between two population, was also calculated using genomics_general (https://github.com/simonhmartin/genomics_general) with a 200-kb sliding window in 20-kb steps. The selected windows simultaneously with top 5% values of the π ratio, F_{ST} and D_{xy} were defined as strong selective sweep regions. In addition, we also independently estimated the F_{ST} between each adhesive egg population (LZL and TL) against each floating egg population (DJKR, HJR, XKL, and YSR) to identify the divergent regions between LZL and TL. Finally, the genes within or overlapping the sweep regions were selected for subsequent gene GO and KEGG pathway enrichment analyses.

3 RESULTS AND DISSCUSSION

3.1 Chromosome-scale genome assembly of topmouth culter

For de novo assembly of the topmouth culter, we integrated data from MGISEQ-T7, PacBio sequencing, and Hi-C platforms, as illustrated in Table S1. After quality control, a total of 132.91 Gb (~100 × depth) of clean reads produced from the MGISEQ-T7 platform were used for genome estimation. The 17-k-mer analysis showed a genome size of 1.2 Gb with a heterozygosity of 0.45% (Table S2). A total of 220.72 Gb (~200 × depth) PacBio sequencing data with a mean length of 22,404 bp were used for assembling the topmouth culter genome.

PacBio clean reads were used to assemble the genome, and finally, a 1.05 Gb

reference genome comprising 262 contigs (> 1 kb) with an N50 length of 17.8 Mb was obtained; the constructed genome is superior to the previously published topmouth culter draft genome (Tables 1 and S3), which had 34,855 contigs covering 1.02 Gb, with an N50 length of 72.24 Kb (Ren et al., 2019). Moreover, the contig N50 length of the topmouth culter genome assembly constructed in this study is higher than that of the published genome assemblies of other related species, for example, the *M. amblycephala* genome had a contig N50 length of 2.40 Mb (Liu et al., 2021) and the *A. nigrocauda* genome had a contig N50 length of 3.12 Mb (Zhang et al., 2020).

3.2 Genome anntotation

Approximately 97.02% of the assembled sequences (1.02 Gb) were anchored onto 24 chromosomes by using 115.56 Gb clean data generated from Hi-C library (Figure 1a, Table S4), consistent with the previously reported karyotype result (Wang et al., 2009). The GC content of the topmouth culter genome was approximately 37.5% (Figures 1b and S2), similar to those of the genomes of other cyprinids (Jian et al., 2021; Xu et al., 2014 Zhang et al., 2020). We observed that approximately 49.79% of the genome assembly accounted for repetitive sequences, with DNA transposons (34.58%) and long terminal repeat retrotransposons (8.11%) being the most abundant transposable elements (Figure 1b, Table S5). In addition, we identified 760,249 SSRs with mono-nucleotide repeat ranked the most (49.9%) (Tables S6-S7). Finally, a total of 26,208 protein-coding genes (Table S8) were identified in the topmouth culter genome by using a combination of de novo strategies and homology-based and RNA-seq-based strategies. The predicted gene models showed similar distribution patterns with those of other

seven fish species in the number and length of CDS, exons, and introns (Figure S3, Table S8). Approximately 99.26% of the genes were successfully annotated by alignment to the public database (Table S9). To evaluate the completeness of the topmouth culter genome assembly, we mapped the MGISEQ-T7 short reads to the assembled genome, which indicated a mapping rate of 99.48% (Table S10). Using BUSCO, the coverage of 3462 highly conserved single-copy Actinopterygii genes was found to be 95.1% and 94.7% for the assembled genome and gene set, respectively (Table S11). Moreover, a high collinearity was observed among the topmouth culter, grass fish and zebrafish genomes (Figures S4-S5). The aforementioned results confirmed that the constructed chromosome-level genome assembly of topmouth culter was of high quality.

3.3 Comparative genomic and evolutionary analyses

To investigate the phylogenetic relationship among topmouth culter and other species and estimate their divergence times, we constructed a phylogenetic tree by using single-copy orthologs (Figure 1c). The results indicated that topmouth culter and *A. nigrocauda* diverged approximately 3.51 MYA after being diverged from *M. amblycephala* at approximately 5.38 MYA. The divergence times for the three Culterinae species were much later than the previously estimated times of divergence of topmouth culter (12.84 MYA) (Ren et al., 2019) and *A. nigrocuda* (8.79 MYA) (Zhang et al., 2020) from *M. amblycephla*, which might be attributed to the higher number of cyprinid species and more calibration time considered in the present study. The divergence time of the three Culterinae species and *C. idellus* was 14.02 MYA in

middle Miocene, coinciding with the time of initiation of the Asian monsoon intensification (Clift et al., 2008), supporting the hypothesis that the burst of diversification in the endemic East Asian cyprinids is related to monsoon activity (Chen, 1998; Chen et al., 2023b; Feng et al., 2022; He et al., 2004).

Gene family evolutionary analysis revealed that 519 gene families were expanded and 267 gene families were contracted in the topmouth culter genome when compared with its most recent common ancestor (Figure 1c). Functional enrichment analysis of the expanded gene families showed that they were significantly enriched in 24 GO terms and 32 KEGG pathways (Table S14), mainly related to proteolysis (GO:0006508, p.adjust = 5.73E-09), DNA integration (GO:0015074, p.adjust = 4.15E-16), motor activity (GO:0003774, p.adjust = 5.95E-07), myosin complex (GO:0016459, p.adjust = 5.95E-07), NOD-like receptor signaling pathway (map04621, p.adjust = 2.07E-41), parathyroid hormone synthesis, secretion and action (map04928, p.adjust = 7.03E-08), and protein digestion and absorption (map04974, p.adjust = 1.80E-06) (Figure S6). The presence of these immune-, nutrition-, and locomotion-related genes in topmouth culter is consistent with its carnivorous habit, unlike the other dietary habits such as herbivorous and phytoplanktivorous in the endemic East Asian cyprinids, indicating that these genes may have a crucial role in species-specific adaptation. The contracted gene families were mainly involved in nucleosome (GO:0000786, p.adjust =2.55E-73), necroptosis (map04217, p.adjust = 1.25E-09), sulfotransferase activity (GO:0008146, p.adjust = 0.007791), and glycosaminoglycan biosynthesis (map00534, p.adjust = 1.81E-05) (Figure S7). The contraction of glycosaminoglycan biosynthesis genes may

be related to the absence of the adhesive layer on the egg envelope of the floating egg.

3.4 Population structure analyses

Understanding the populationstructure of topmouth culter holds great importance for conservation and genetic breeding studies. Therefore, we re-sequenced 158 topmouth culter individuals from six populations, representing three geographical groups (Amur River basin, Yangtze River basin, and Peral River basin) and two egg type variations (floating and adhesive egg) (Figure 2a). An average size of 15.3 Gb (~14.84×) 2 × 150 paired data per individual was generated with an average mapping rate and coverage of 99.26% and 96.69%, respectively (Table S15). After SNP calling and filtering, a total of 7,276,044 and 1,587,880 high-quality SNPs and InDels, respectively, were detected.

Admixture analysis revealed three genetically distinct clusters that were strongly partitioned by geographic proximity (Figure 2c). XKL belongs to Amur River basin and HJR belongs to Peral River basin were successively separated from the populations in Yangtze River basin when the ancestry components (*K*) increased from 2 to 3. When the best-support for $K = 4$ (Figure S8), LZL and YSR populations in the Yangtze River basin clustered together, whereas DJKR and TL populations formed one cluster despite being located at a long distance. Considering that the DJKR, situated in the upstream reaches of a Yangtze River tributary, was constructed several decades ago (Sun et al., 2021), the genetic similarity between DJKR and TL may be attributed to a shared ancestral polymorphism. Notably, the XKL population showed no admixture when the K value increased to 6, suggesting its greater genetic distance from the other populations, consistent with the findings of previous studies (Sun et al., 2021; Xiong et

al., 2019). Interestingly, the two adhesive populations LZL and TL showed some degree of admixture, indicating potential gene flow or parallel adaptive divergence. The NJ tree (Figure 2b) and PCA (Figure 2d) recapitulated these groupings. Additionally, the position of the outgroup species *A. nigrocauda* in the NJ tree suggested that *A. nigrocauda* was closer to the topmouth culter population in YSR, consistent with its sympatric distribution in the upper reaches of Yangtze River.

The PSMC analysis revealed two rounds of population declines, which corroborated well with the Tibetan Plateau uplifting events (Figure 2e). The peak of effective population size (Ne) of the six topmouth culter populations was nearly 3.5 MYA, followed by a sharp decline, coinciding with two intense uplift phases, which are, Qingzang (3.6–1.7 MYA) and Kunhang (1.1–0.6 MYA) movements in the third tectonic Tibetan Plateau uplift phase. The second population decline occurred with the beginning of Gonghe Movement (~0.15 MYA) (An et al., 2001; Li & Fang, 1999). This demographic pattern may be attributed to the remarkable changes in geology and climate during Tibetan Plateau uplifting, which may be unfavorable to the topmouth culter population.

3.5 Genetic diversity and linkage disequilibrium

To evaluate the divergence degree among the three geographical populations of the topmouth culter from six locations, we firstly calculated the genetic diversity indexes and their pairwise population differentiation coefficient F_{ST}. The observed and expected heterozygosity values were similar among populations, with H_O ranged from 0.27 to 0.30 and H_E ranged from 0.30 to 0.31 (Table S12). Nucleotide diversity (π) within the

population exhibited the highest value in TL (1.94×10^{-3}), followed by LZL (1.93×10^{-3}), YSR (1.92×10^{-3}), DJKR (1.88×10^{-3}) and HJR (1.78×10^{-3}), while the lowest value was observed in XKL (1.53×10^{-3}) (Table S12). This pattern aligns with the presence of egg type variation in the Yangtze River basin. The comparison of F_{ST} also illustrated that the XKL population in the Amur basin was more distant from the populations in Yangtze River basin than the HJR population in the Peral River basin, which was also supported by the results of population structure (Figure 3a; Table S13). The LD decay rates of the six populations varied markedly, with the highest LD level was found in XKL, followed by that in HJR, indicating a stronger bottleneck or the founder effect (Bray et al., 2010) (Figure 3b). Overall, these results suggest that the genetic differentiation among the six topmouth culter populations primarily resulted from geographic isolation, and XKL population with the lowest genetic diversity requires enhanced conservation efforts.

3.6 Selection signatures of egg type variation in topmouth culter

Semi-buoyant eggs of topmouth culter undergo substantial hydration, whereas adhesive eggs possess a unique adhesive layer on their envelope, which is responsible for specific adaptations to spawning environments (Chen et al., 2022). The intraspecific variation of egg type in different topmouth culter populations suggested a divergent selection, which may be a strong evolutionary force driving population differentiation. Therefore, we conducted selective sweeps detection to find outlier SNPs or diverged regions between the adhesive and floating egg populations.

Conjoint analysis of π ratios, F_{ST} and D_{xy} (both top 5%) identified divergent genomic regions containing 72 and 94 genes for the adhesive group (LZL and TL) and floating group (DJKR, HJR, XKL, and YSR), respectively (Figure 4a; Table S16). Specifically, GO and KEGG enrichment analysis revealed a significant number of genes were represented in the pathways of regulation of actin cytoskeleton and lipid metabolic (Figure 4b; Figure S9), consistent with the processes of fertilization and egg activation during topmouth culter embryogenesis (Chen et al., 2023). We found high levels of heterogeneous genomic divergence between the two phenotypic differentiation populations scattered across the genome and identified selection signals spanning a set of candidate genes overlapped with the key pathways of hydration and adhesiveness in the topmouth culter eggs (Chen et al., 2022). For example, Zinc finger protein (ZFP) in the zinc metalloproteinase pathway might play a role in the yolk protein degradation. Voltage-dependent calcium channel (CACN) might be involved in egg envelope permeability transition pore. Procollagen galactosyltransferase (COLGALT), collagen alpha-4(VI) chain (COL6A4), COL6A6, fibronectin type III domain-containing protein 5 (FNDC5) and integrin alpha-X (ITGAX) as the crosslinks of microfilament-associated proteins might contribute to the adhesiveness of adhesive eggs (Tang 2020; Whittaker & Hynes, 2002). We also independently compared each adhesive egg population against each floating egg population. The pairwise comparison revealed similar genomic divergence (genomic islands) between the two adhesive populations (Figure S10) and identified many overlapped genes that can be indicative of local adaptation responses to different hydrological environments (Table S17).

Taken together, we believe that these candidate genes may be valuable for further functional characterization.

4 CONCLUSIONS

The present study reports a chromosome-scale genome assembly for topmouth culter. In this study, the genetic relationship of six topmouth culter populations and the genomic variation between adhesive and semi-buoyant egg phenotypes based on whole-genome resequencing data of 158 individuals were explored. The topmouth culter genome constructed in this study is of high quality, with a contig N50 length of 17.8 Mb, and shows high completeness (BUSCO score, 96.7%). Comparative genomic and evolutionary analyses revealed the divergence time and genetic variation of topmouth culter with other endemic East Asian cyprinids. The population-level genetic analysis revealed distinctive geographical groups and a significantly declined genetic diversity in the XKL population. The study also analyzed signatures of selection toward egg type variation in the adhesive and floating populations. Genomic data obtained in the present study can serve as a valuable resource for further studies on evolution, genetic breeding, and conservation of topmouth culter and other endemic East Asian cyprinids.

ACKNOWLEDGEMENTS

We would like to thank Prof. Dongli Qin, Heilongjiang River Fisheries Research Institute, Chinese Academy of Fishery Sciences, for assistance in sample collection.

This research was supported by the National Natural Science Foundation of China (32102797) and Natural Science Foundation of Fujian Province of China (2022J011136).

CONFLICT OF INTEREST

The authors declare no competing interests.

AUTHOR CONTRIBUTIONS

H.J. and Y.Q. conceived and led the study; S.H., L.Y. and G.X. designed this project and research aspects; Z.Z., M.M., and Y.D. performed sample collection and data analyses. All authors were involved in the writing of the paper and approved the final manuscript.

DATA AVAIL ABILIT Y S TATEMENT

All the genomic reads (MGISEQ-T7, PacBio and Hi-C sequencing data) including transcriptome data and resequencing data generated in this study have been deposited at the China National Center for Bioinformation (CNCB, https://www.cncb.ac.cn/) under the accession no. PRJCA011991. Genome assembly is under accession No. GWHBOSX00000000. All other study data are included in the article and/or supporting information.

REFERENCES

Alexander, D. H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *Bmc Bioinformatics, 12*(1),*246*. https://doi.org/10.1186/1471-2105-12-246

An, Z. S., Kutzbach, J. E., Prell, W. L., & Porter, S. C. (2001). Evolution of Asian monsoons and phased uplift of the Himalayan Tibetan plateau since Late Miocene times. *Nature, 411*(6833), 62-66. https://doi.org/10.1038/35075035

Bray, S. M., Mulle, J. G., Dodd, A. F., Pulver, A. E., Wooding, S., & Warren, S. T. (2010). Signatures of founder effects, admixture, and selection in the Ashkenazi Jewish population. *Proceedings of the National Academy of Sciences of the United States of America, 107*(37), 16222-16227. https://doi.org/10.1073/pnas.1004381107

Burton, J. N., Adey, A., Patwardhan, R. P., Qiu, R. L., Kitzman, J. O., & Shendure, J. (2013). Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nature Biotechnology, 31*(12), 1119-1125. https://doi.org/10.1038/nbt.2727

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution, 17*(4), 540-552. https://doi.org/10.1093/oxfordjournals.molbev.a026334

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience, 4*(1), s13742–s14015. https://doi.org/10.1186/s13742-015-0047-8

Chen, F., Smith, C., Wang, Y. K., He, J., Xia, W. L., Xue, G., Chen, J., & Xie, P. (2021). The Evolution of Alternative Buoyancy Mechanisms in Freshwater Fish Eggs. *Frontiers in Ecology and Evolution, 9,* 736718. https://doi.org/10.3389/fevo.2021.736718

Chen, F., Wang, Y. K., He, J., Chen, L., Xue, G., Zhao, Y., Peng, Y. H., Smith, C., Zhang, J., Chen, J., & Xie, P. (2022). Molecular Mechanisms of Spawning Habits for the Adaptive Radiation of Endemic East Asian Cyprinid Fishes. *Research, 9827986*. https://doi.org/10.34133/2022/9827986

Chen, F., Wang, Y. K., He, J., Smith, C., Xue, G., Zhao, Y., Peng, Y. H., Zhang, J., Liu, J. R., Chen, J., & Xie, P. (2023a). Alternative signal pathways underly fertilization and egg activation in a fish with contrasting modes of spawning. *BMC Genomics, 24,* 167. https://doi.org/10.1186/s12864-023-09244-1

Chen, F., Xue, G., Wang, Y. K., Zhang, H. C., Clift, P. D., Xing, Y. W., He, J., Albert, J. S., Chen, J., & Xie, P. (2023b). Evolution of the Yangtze River and its biodiversity. *Innovation, 4(3),* 100417. https://doi.org/10.1016/j.xinn.2023.100417

Chen, Y. Y. (1998). Fauna Sinica, Osteichthys: Cypriniformes (Part II). Science Press, Beijing, China (in Chinese).

Cheng, P., Yu, D., Tang, Q., Yang, J., Chen, Y., & Liu, H. (2022). Macro-evolutionary patterns of East Asian opsariichthyin-xenocyprinin-cultrin fishes related to the formation of river and river-lake environments under monsoon climate. *Water Biology and Security, 1*(2). https://doi.org/10.1016/j.watbs.2022.100036

Chin, C.-S., Peluso, P., Sedlazeck, F. J., Nattestad, M., Concepcion, G. T., Clum, A., Dunn, C., O'Malley, R., Figueroa-Balderas, R., MoralesCruz, A., Cramer, G. R., Delledonne, M., Luo, C., Ecker, J. R., Cantu, D., Rank, D. R., & Schatz, M. C. (2016). Phased Diploid Genome Assembly with Single Molecule Real-Time Sequencing. *Nature Methods, 13*(12), 1050. https://doi.org/10.1038/nmeth.4035

Clift, P. D., Hodges, K. V., Heslop, D., Hannigan, R., Van Long, H., Calves, G. (2008). Correlation of Himalayan exhumation rates and Asian monsoon intensity. *Nature geoscience, 1*, 875-880. https://doi.org/10.1038/ngeo351

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M.A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics, 27*, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

De Bie, T., Cristianini, N., Demuth, J. P., & Hahn, M. W. (2006). CAFÉ: a computational tool for the study of gene family evolution. *Bioinformatics, 22*(10), 1269-1271. https://doi.org/10.1093/bioinformatics/btl097

Feng, C. G., Wang, K., Xu, W. J., Yang, L. D., Wanghe, K. Y., Sun, N., Wu, B. S., Wu, F. X., Yang, L., Qiu, Q., Gan, X. N., Chen, Y. Y., & He, S. P. (2022) Monsoon boosted radiation of the endemic East Asian carps. *Science China-Life Science*, *65*. https://doi.org/10.1007/s11427-022-2141-1

Haas, B. J., Delcher, A. L., Mount, S. M., Wortman, J. R., Smith, R. K., Hannick, L. I., Maiti, R., Ronning, C. M., Rusch, D. B., Town, C. D., Salzberg, S. L., & White, O. (2003). Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Research, 31*(19), 5654-5666. https://doi.org/10.1093/nar/gkg770

Haas, B. J., Salzberg, S. L., Wei, Z., Pertea, M., Allen, J. E., Orvis, J., White, O., Buell, C. R., & Wortman, J. R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biology, 9*(1), R7. https://doi.org/10.1186/gb-2008-9-1-r7

He, S. P., Liu, H. Z., Chen, Y. Y., Kuwahara, M., Nakajima, T., & Zhong, Y. (2004). Molecular phylogenetic relationships of Eastern Asian Cyprinidae (Pisces: Cypriniformes) inferred from cytochrome b sequences. *Science in China Series C-Life Sciences, 47*(2), 130-138. https://doi.org/10.1360/03yc0034

Hu, J., Fan, J. P., Sun, Z. Y., & Liu, S. L. (2020). NextPolish: A fast and efficient genome polishing tool for long-read assembly. *Bioinformatics*, *36*(7), 2253–2255. https://doi.org/10.1093/bioinformatics/btz891

Jessen, J. R. (2015). Recent advances in the study of zebrafish extracellular matrix proteins. *Developmental Biology*, *401*(1), 110-121. https://doi:10.1016/j.ydbio.2014.12.022

Jian, J. B., Yang, L. D., Gan, X. N., Wu, B., Gao, L., Zeng, H. H., Wang, X. Z., Liang, Z. Q., Wang, Y., Fang, L H., Li, J., Jiang, S. J., Du, K., Fu, B. D., Bai, M. Z., Chen, M., Fang, X. D., Liu, H. Z., & He, S. P. (2021). Whole genome sequencing of silver carp (*Hypophthalmichthys molitrix*) and bighead carp (*Hypophthalmichthys nobilis*) provide novel insights into their evolution and speciation. *Molecular Ecology Resource, 21*(3), 912-923. https://doi.org/10.1111/1755-0998.13297

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution, 30*(4), 772-780. https://doi.org/10.1093/molbev/mst010

Kawaguchi, M., Yasumasu, S., Shimizu, A., Kudo, N., Sano, K., Iuchi, I., & Nishida, M. (2013). Adaptive evolution of fish hatching enzyme: one amino acid substitution results in differential salt dependency of the enzyme. *Journal of Experimental Biology, 216*(Pt 9), 1609-1615. https://doi.org/10.1242/jeb.069716

570 Keilwagen, J., Wenk, M., Erickson, J. L., Schattat, M. H., Grau, J., & Hartung, F. (2016). Using
571 intron position conservation for homology-based gene prediction. *Nucleic Acids Research,*
572 *44*(9), e89. https://doi.org/10.1093/nar/gkw092
573 Kumar, S., Stecher, G., Suleski, M., & Hedges, S. B. (2017). TimeTree: A Resource for Timelines,
574 Timetrees, and Divergence Times. *Molecular Biology and Evolution, 34*(7), 1812-1819.
575 https://doi.org/10.1093/molbev/msx116
576 Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature*
577 *Methods, 9*(4), 357-359. https://doi.org/10.1038/nmeth.1923
578 Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler
579 transform. *Bioinformatics, 25*(14), 1754-1760.
580 https://doi.org/10.1093/bioinformatics/btp324
581 Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome
582 sequences. *Nature, 475*(7357), 493-U484. https://doi.org/10.1038/nature10231
583 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., & Homer, N., Marth, G., Abecasis, G.,
584 Durbin, R., & Genome Project Data Processing, S (2009). The Sequence Alignment/Map
585 format and SAMtools. *Bioinformatics, 25*(16), 2078-2079.
586 https://doi.org/10.1093/bioinformatics/btp352
587 Li, J. J., & Fang, X. M. (1999). Uplift of the Tibetan Plateau and environmental changes. *Chinese*
588 *Science Bulletin, 44*(23), 2117-2124. https://doi.org/10.1007/BF03182692
589 Li, L., Stoeckert, C. J., & Roos, D. S. (2003). OrthoMCL: Identification of ortholog groups for
590 eukaryotic genomes. *Genome Research, 13*(9), 2178-2189.
591 https://doi.org/10.1101/gr.1224503

592 Liu, H., Chen, C. H., Lv, M. L., Liu, N., Hu, Y. F., Zhang, H. L., Enbody, E. D., Gao, Z. X.,
593 Andersson, L., & Wang, W. M. (2021). A chromosome-level assembly of blunt snout bream
594 (*Megalobrama amblycephala*) reveals an expansion of olfactory receptor genes in
595 freshwater fish. *Molecular Biology and Evolution*.
596 https://doi.org/10.1093/molbev/msab152
597 McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K.,
598 Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis
599 Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data.
600 *Genome Research, 20*(9), 1297-1303. https://doi.org/10.1101/gr.107524.110
601 Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., &
602 Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic
603 Inference in the Genomic Era. *Molecular Biology and Evolution, 37*(5), 1530-1534.
604 https://doi.org/10.1093/molbev/msaa015
605 Nelson, J. S., Grande, T. C., & Wilson, M. V. (2016). *Fishes of the World*: John Wiley & Sons.
606 Ou, S. J., Su, W. J., Liao, Y., Chougule, K., Agda, J. R. A., Hellinga, A. J., Lugo, C. S. B., Elliott, T.
607 A., Ware, D., Peterson, T., Jiang, N., & Hufford, M. B. (2019). Benchmarking transposable
608 element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome*
609 *Biology, 20*(1). 275. https://doi.org/10.1186/s13059-019-1905-y
610 Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *Plos*
611 *Genetics, 2*(12), 2074-2093. https://doi.org/10.1371/journal.pgen.0020190
612 Qi, P.Z., Qin, J. H., & Xie, C. X. (2015). Determination of genetic diversity of wild and cultured
613 topmouth culter (*Culter alburnus*) inhabiting China using mitochondrial DNA and

microsatellites. *Biochemical Systematics and Ecology, 61*, 232-239. https://doi.org/10.1016/j.bse.2015.06.023

Ren, L., Li, W.H., Qin, Q. B., Dai, H., Han, F. M., Xiao, J., Gao, X., Cui J. L., Wu, C., Yan, X. j., Wang, G. L., Liu, G. M., Liu, J., Li, J. M., Wan, Z., Yang, C. H., Zhang, C., Tao. M., Wang, J., Luo, K. K., Wang, S., Hu, F. Z., Zhao, R. R., Li, X. M., Liu, M., Zheng, H. K., Zhou, R., Shu, Y. Q., Wang, Y. D., Liu, Q. F., Tang, C. C., Duan, W., & Liu, S. J. (2019). The subgenomes show asymmetric expression of alleles in hybrid lineages of *Megalobrama amblycephala* × *Culter alburnus*. *Genome Research, 29*(11), 1805-1815. https://doi.org/10.1101/gr.249805.119

Servant, N., Varoquaux, N., Lajoie, B. R., Viara, E., Chen, C. J., Vert, J. P., Heard, E., Dekker, J., & Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biology, 16*, 259. https://doi.org/10.1186/s13059-015-0831-x

Silvestro, D., & Michalak, I. (2012). raxmlGUI: a graphical front-end for RAxML. *Organisms Diversity & Evolution, 12*(4), 335-337. https://doi.org/10.1007/s13127-011-0056-0

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics, 31*(19), 3210. https://doi.org/10.1093/bioinformatics/btv351

Sun, N., Zhu, D. M., Li, Q., Wang, G. Y., Chen, J., Zheng, F. F., Li, P., & Sun, Y. H. (2021). Genetic diversity analysis of Topmouth Culter (*Culter alburnus*) based on microsatellites and D-loop sequences. *Environmental Biology of Fishes, 104*(3), 213-228. https://doi.org/10.1007/s10641-021-01062-2

Tan, M., & Armbruster, J. W. (2018). Phylogenetic classification of extant genera of fishes of the

order Cypriniformes (Teleostei: Ostariophysi). *Zootaxa, 4476*(1), 6-39. https://doi.org/10.11646/zootaxa.4476.1.4

Tang, V. W. (2020). Collagen, stiffness, and adhesion: the evolutionary basis of vertebrate mechanobiology. *Molecular Biology of the Cell, 31*(17), 1823-1834. https://doi.org/10.1091/mbc.E19-12-0709

Thiel, T., Michalek, W., Varshney, R. K., & Graner, A. (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical & Applied Genetics, 106*(3), 411-422. https://doi.org/10.1007/s00122-002-1031-0

Wang, X. Z., Li, J. B., & He, S. P. (2007). Molecular evidence for the monophyly of East Asian groups of Cyprinidae (Teleostei: Cypriniformes) derived from the nuclear recombination activating gene 2 sequences. *Molecular Phylogenetics and Evolution, 42*(1), 157-170. https://doi.org/10.1016/j.ympev.2006.06.014

Wang, Y. P., Lu, Y., Zhang, Y., Ning, Z. M., Li, Y., Zhao, Q., Lu, H. Y., Huang, R., Xia, X. Q., Feng, Q., Liang, X. F., Liu, K. Y., Zhang, L., Lu, T. T., Huang, T., Fan, D. L., Weng, Q. J., Zhu, C. R., Lu, Y. Q., … Zhu, Z. Y. (2015). The draft genome of the grass carp (*Ctenopharyngodon idellus*) provides insights into its evolution and vegetarian adaptation. *Nature Genetics, 47*(6), 625–631. https://doi.org/10.1038/ng.3280

Whittaker, C. A., & Hynes, R. O. (2002). Distribution and evolution of von Willebrand/integrin A domains: widely dispersed domains with roles in cell adhesion and elsewhere. *Molecular Biology of the Cell, 13*(10), 3369-3387. https://doi.org/10.1091/mbc.e02-05-0259

Xiong, Y., Li, W., Yuan, J., Zhang, T., Li, Z., Xiao, W., & Liu, J. (2019). Genetic structure and

demographic histories of two sympatric Culter species in eastern China. *Journal of Oceanology and Limnology, 38*(2), 408-426. https://doi.org/10.1007/s00343-019-9036-6

Xu, P., Xu, J., Liu, G. J., Chen, L., Zhou, Z. X., Peng, W. Z., Jiang, Y. L., Zhao, Z. X., Jia, Z. Y., Sun, Y. H., Wu, Y. D., Chen, B. H., Pu, F., Feng, J. X., Luo, J., Chai, J., Zhang, H, J., Wang, H., Dong, C. J., Jiang, W. K., & Sun X. W. (2019). The allotetraploid origin and asymmetrical genome evolution of the common carp *Cyprinus carpio*. *Nature Communcations, 10*(1), 4625. https://doi.org/10.1038/s41467-019-12644-1

Xu, P., Zhang, X. F., Wang, X. M., Li, J. T., Liu, G. M., Kuang, Y. Y., Xu, J., Zheng, X. H., Ren, L. F., Wang, G. L., Zhang, Y., Huo, L. H., Zhao, Z. X., Cao, D. C., Lu, C. Y., Li, C., Zhou, Y., Liu, Z. J., Fan, Z. H., ... Sun, X. W. (2014). Genome sequence and genetic diversity of the common carp, *Cyprinus carpio*. *Nature Genetics, 46*(11), 1212-1219. https://doi.org/10.1038/ng.3098

Yagi, M., Kosugi, S., Hirakawa, H., Ohmiya, A., Tanase, K., Harada, T., Kishimoto, K., Nakayama, M., Ichimura, K., Onozaki, T., Yamaguchi, Y., Sasaki, N., Miyahara, T., Nishizaki, Y., Ozeki, Y., Nakamura, N., Suzuki, T., Tanaka, Y., Sato, S., Shirasawa, K., Isobe, S., Miyamura, Y., Watanabe, A., Nakayama, S., Kishida, Y., Kohara, M & Tabata, S. (2014). Sequence Analysis of the Genome of Carnation (*Dianthus caryophyllus* L.). *Dna Research, 21*(3), 231-241. https://doi.org/10.1093/dnares/dst053

Yang, Z. H. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution, 24*(8), 1586-1591. https://doi.org/10.1093/molbev/msm088

Zhang, H. H., Xu, M. R., Wang, P. L., Zhu, Z. G., Nie, C. F., Xiong, X. M., Wang, L., Xie, Z. Z., Wen, X., Zeng, Q. X., Zhang, X. G., & Dai, F. Y. (2020). High-quality genome assembly

and transcriptome of *Ancherythroculter nigrocauda*, an endemic Chinese cyprinid species. *Molecular Ecology Resources, 20*(4), 882-891. https://doi.org/10.1111/1755-0998.13158

Zhu, K. Y., Du, P. X., Xiong, J. X., Ren, X. Y., Sun, C., Tao, Y. C., Ding, Y., Xu, Y. R., Meng, H. L., Wang, C. C., & Wen, S. Q. (2021). Comparative Performance of the MGISEQ-2000 and Illumina X-Ten Sequencing Platforms for Paleogenomics. *Frontiers in Genetics, 12*, 745508. https://doi.org/10.3389/fgene.2021.745508

686 # Tables and Figures

687 ## Tables

688 **TABLE 1** Comparison of our genome assembly of topmouth culter with previous study

Assembly	**This study**	**GCA_009869775.1[a]**
Assembly approach	MGISEQ-T7, Pacbio, HiC	Illumina, PacBio
Assembled genome size (Gb)	1.05	1.02
Contig number	262	34,855
Contig length (bp)	1,053,229,386	991,157,727
Contig N50 (bp)	17,799,895	72,243
Contig N90 (bp)	2,878,033	14,789
Contig maximum (bp)	44,146,744	614,399
GC (%)	37.50	37.36
Gap number	125	29,167

689 [a]Previously published version (Ren et al., 2019).

Figure legends

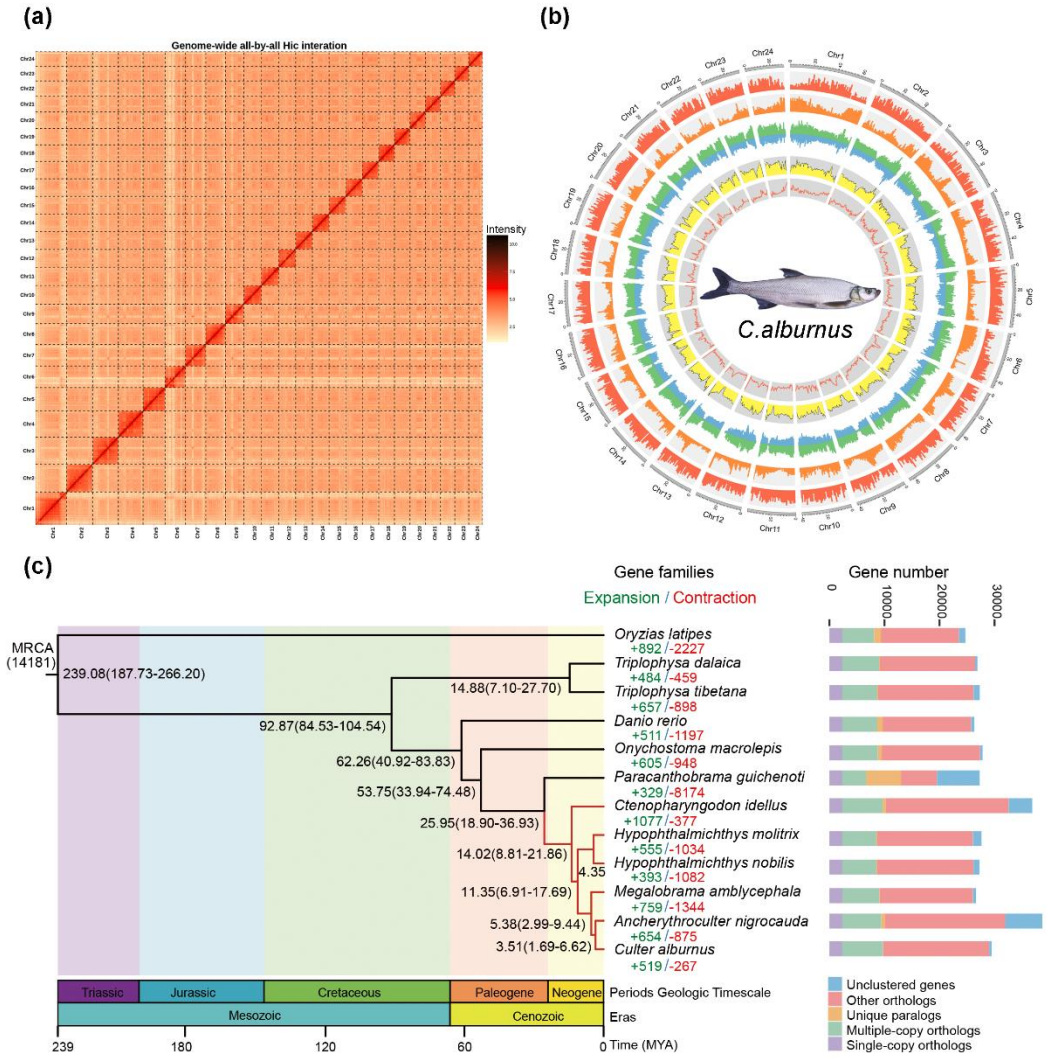


Figure 1 Genome features and phylogenetic and evolutionary analyses. (a) Interaction matrix across the topmouth culter genome; blocks with higher color intensity indicate stronger contacts. (b) Circos graph (from outside to inside) representing the gene density, all repeat sequence density, SNP (green) and InDel (blue) density, total genetic diversity (π), and the GC content distribution across the chromosomes of the genome with 1-Mb sliding window size. (c) Phylogenetic tree based on 2106 single-copy orthologs and distribution of homologous genes of the 12 species. The numbers near the ancestral nodes indicate the estimated divergence time (MYA, million years ago),

with the 95% confidence intervals in parentheses. Endemic East Asian cyprinid lineages are indicated in red. Expansion and contraction of gene families are represented as green and red numerical values, respectively. The stacked-column plot illustrates the distribution of unique genes, single-/multiple-copy genes, other, and unclustered genes.

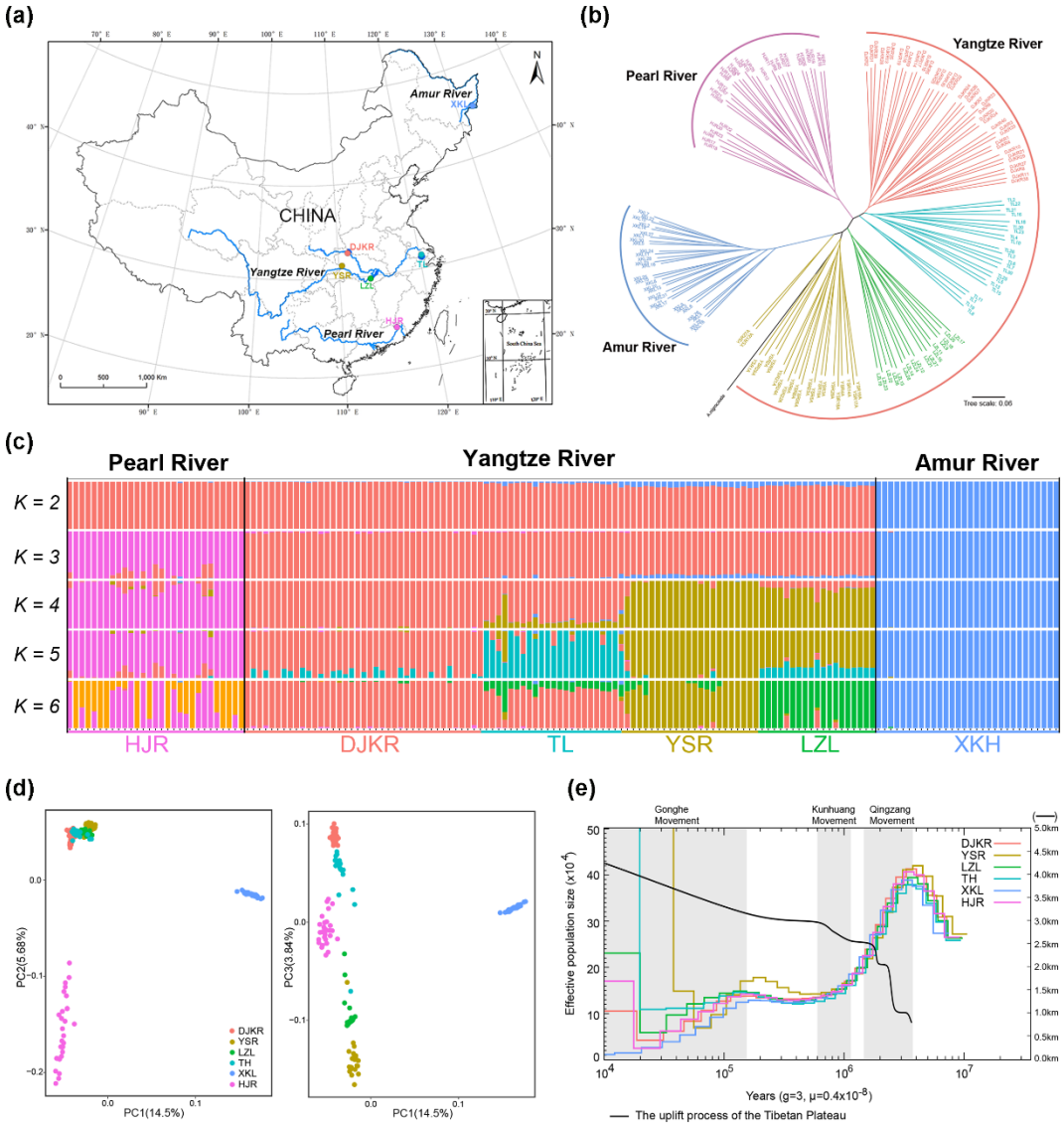
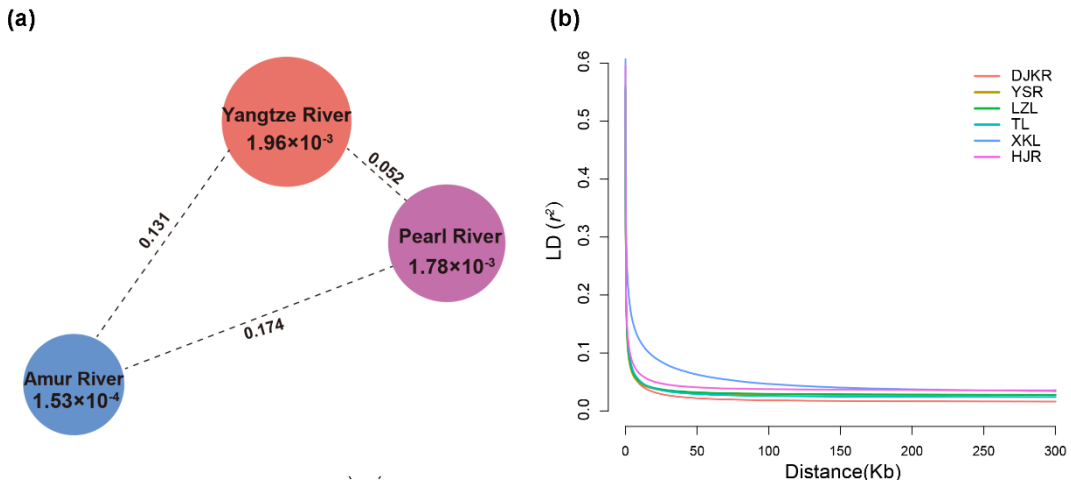


Figure 2 Population structure analysis and demographic history of different topmouth culter geographical populations. (a) Geographic locations of the six topmouth culter populations. Circles with different colors represent different geographic sites. (b) Phylogenetic tree inferred from whole-genome SNPs. (c) Genetic structure of topmouth culter with different ancestry kinships (K = 2 to 6). Each bar represents an individual, and different colors represent the proportion contributed by that ancestral population. Different geographical populations are indicated along the bottom X-axis with different colors, as indicated in (a). (d) PCA plots of the first three components of the 158

topmouth culter individuals. (e) Demographic histories constructed using the PSMC model. The time range of three rounds of intense uplift (Qingzang, Kunhuang and Gonghe Movement) is shaded in gray. The black curve shows the Tibetan Plateau uplift event, and the right Y-axis indicates the height above sea level.



717

718 **Figure 3** Population diversity analysis. (a) genetic divergence across three basins

719 studied. The value in each circle represented nucleotide diversity (π) for this group, and

720 the pairwise genetic divergence (F_{ST}) is indicated on the line linked two sites. (b)

721 Linkage disequilibrium distance analysis.

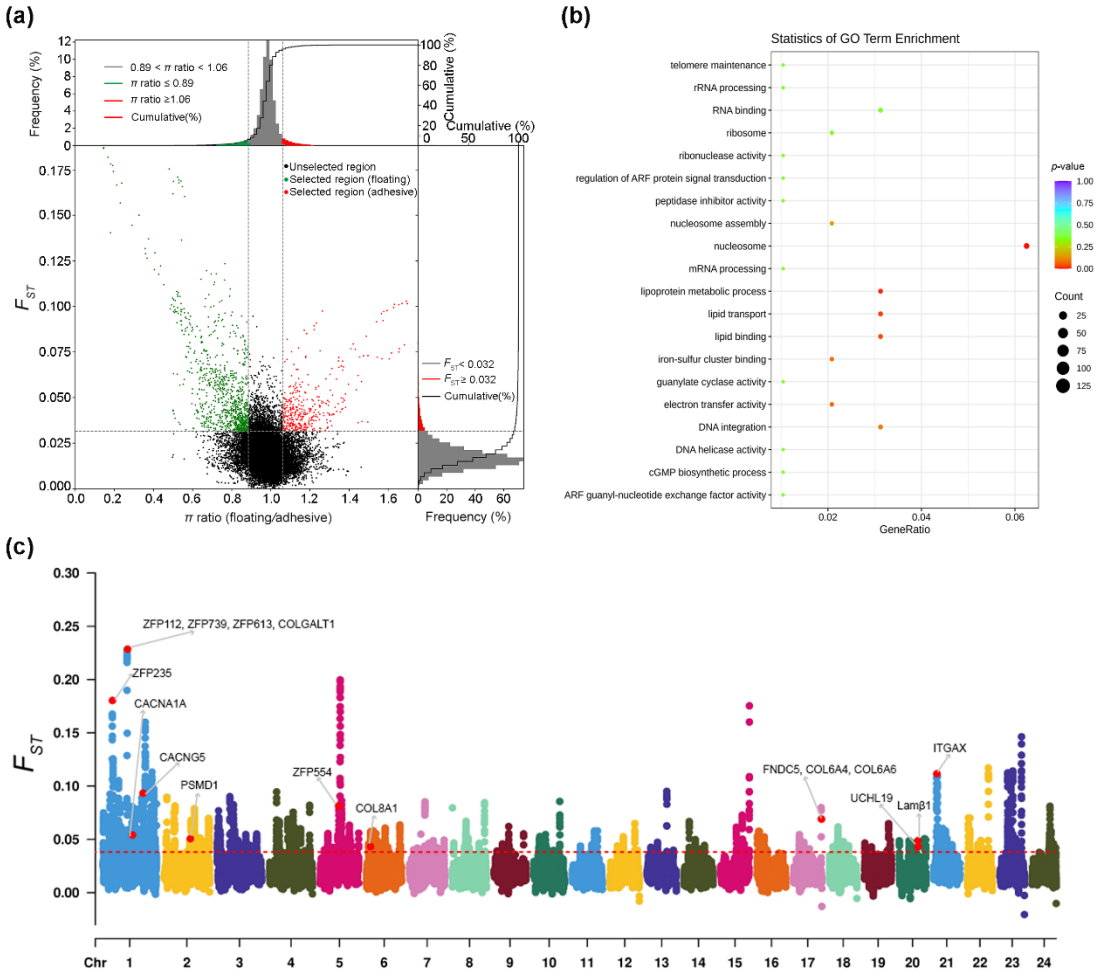


Figure 4 Identification of divergent regions in the floating and adhesive egg topmouth culter populations. (a) Distribution of π ratios (floating/adhesive) and F_{ST} values, which were calculated in 200-kb windows sliding in 20-kb steps. Green points in the upper left panel are the selective sweep regions for the floating populations, whereas red points in the upper right panel are the selective sweep regions for the adhesive egg populations. Vertical and horizontal dashed lines correspond to the 5% tails of the empirical π ratio (0.89 and 1.06) and F_{ST} (0.032) distributions, respectively. (b) Top 20 GO terms of the divergent genes between the floating and adhesive egg populations. (c) Manhattan plot of the highly divergent genomic regions and overlapping selective signals. Candidate genes associated with egg type variation are highlighted in red dots.

733 The dashed line indicates the threshold for selected regions (F_{ST} =0.032). Chr.,

734 chromosome.